MAY 10 2022 10:31 AM THOUGHTS ON PROTEINS AND HUMAN NATURE

ZULFIKAR MOINUDDIN AHMED
ZULFIKAR.AHMED@GMAIL.COM

The difference in the genetic code of one man and an another lies snugly constrained in 0.1% of the genome. The proteins in one man's body to another is similarly restricted. The proteome of one man and that of an another is unlikely to contain radically different elements at all. This is the issue of the universality of the proteome. Biotechnology's fondness for the excitement of analytical chemistry, of separations and mass spectrometry today takes some attention from the problem at hand, how to put the proteome in a universal canonical context. This is the next frontier of advancement for proteomics, to have some handle for protein level expression.

A large class of problems in biology and medicine will benefit when there is a canonical protein expression that is the output of measurement.

I applied today to two jobs in proteomics at Tegmine Therapeutics. What occurs to be as valuable is that universal proteome and canonical output from separations-mass spectrometry will eventually give us applications in universal human psychology as well.

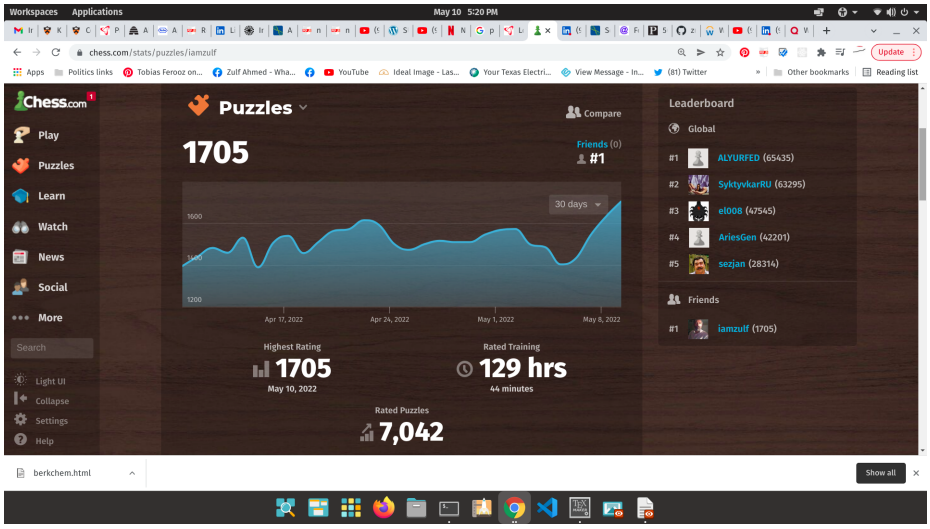
Proteomics is artificially limited today to a small set of applications but with some changes it can have applications to universal human psychology as well. And this is the most important and ubiquitous of application areas.

I realise that without much attention and use for a decade, my resume is not in good order. I want to focus on some of the strengths of my actual experience. I have had significant experience in Finance, Biotechnology and Technology all of them senior and nontrivial positions and all of them requiring high level nontrivial skills and professional competence. In addition from 2008 I have been working on fundamental science issues with many innovative breakthroughs.

I think it is worthwhile to highlight three positions that have been highly nontrivial. First is Fixed Income Finance Quant at Lehman, 1995-1996 where I began coding and learned the ropes of pricing and hedging derivatives. Second is a position as Senior Software Architect 1997-1998 with Ecosystems where I built the entire backend of an object oriented document management database. Third is a position with Biospect/Predicant 2003-2006 where I learned to work on separations-mass spectrometry data for signal processing and statistical learning methods for biomarker discovery. These are the positions that honed my complete palette of nontrivial skills and competence in work that I can do.

Date: May 13, 2022.

1. My Chess Rating Crosses 1700



This is quite exciting because this shows that progress is possible via habituation in chess. With practice, I graduated to 1700 for the first time. As the reader will recall I am not a lifelong chess player, so my progress is a matter of motivated habituation, one of the most central concepts in psychology of universal human behaviour.

2. Rejection For Genentech Senior Scientist Proteomics Job

I am slightly depressed about getting a rejection from Genentech Senior Scientist Proteomics job. I really felt it was a mutual opportunity as I am surely capable of doing good work at the job since not only do I have *experience* doing this sort of work from my Predicant/Biospect years but in the intervening years I have come up with a solution to the protein folding problem and was excited to do some proteomics work again.

Most probably this will be a minor event in my life and that of Genentech. But it is poignant nonetheless for by my successful work in four-sphere theory which includes a replacement for James Clerk Maxwell's laws of electromagnetism as well as Erwin Schroedinger's I know with certainty that history of science will treat me as an immortal genius. There are no laws of nature that ensure that an immortal genius will be able to secure employment at a reasonable job at Genentech, however, as empirical evidence shows.

Historians of science in the future might examine this phenomenon as perhaps common occurrence in the industry where immortal scientific advances are not valued in industry as much as more pragmatic skills and industry's habitual understanding of skillsets needed for industrial goals.

Now I do have some ideas about far-reaching advances in proteomics, the most important of which are derived from my ideas of integrative analysis for separations mass spectrometry from my Predicant/Biospect years with John Stults and others. The main idea is that for every apparatus of separations mass spectrometry the results ought to be a relatively low dimensional vector of protein identity (at least

probabilistic) and their total concentrations. A serious effort here ought to transform the entire industry. The information content of separations mass spectra have never been seriously considered independently enough to understand that a great deal of post-measurement analysis suffer from integration of platform variables and pure scientific content resolved. The problem is not trivial; it is challenging, but it can have a canonical solution that advances the entire industry.

3. Using Dictionary And Convex Optimisation For The Problem

I have extensive knowledge of using convex optimisation techniques for signal modeling of separations mass spectra. Basis pursuit with L^1 penalties is a fairly efficient path for the problem. Suppose B is the vector of ionized proteins with fragments at muliple m/z regions. We can find an optimal fit to the signal with

$$Bx = y$$

with y our signal with optimising objective function

$$\|Bx - y\|_2 + \lambda\|x\|_1$$

There is quite a bit of literature as well as python package cvxopt that allows us to solve these problems [4].

The domain knowledge is B that depends on the separations-mass spectrometry instruments. With a sufficiently large library of basis functions efficiently the technique will yield good estimates of concentration of proteins in biological fluids.

What I am proposing here I am quite confident will work because I worked on this problem for years at Biospect/Predicant and I know this approach yields results but more importantly I am confident that the *whole proteomics field* will advance substantially if it adopts this path.

There are both theoretical and pragmatic methods for tabulating the overdetermined basis matrix B and this is the crux of the scientific problem. I had determined some theoretical solutions 2003–2006.

Ideally one would like to have a uniquely specified B for each instrument configuration and package the software with the instrument where the result of the measurement is just the list of proteins and their concentrations. This would then enable simpler statistical techniques to apply for all possible applications of high throughput proteomics permanently reducing a large number of technical problems of analysing full mass spectra. In a sense the mass spectrum does not contain any interesting information about the sample beyond the identity and concentrations of the proteins.

4. Fast Convex Optimisation Is A Technology

There is a book called *Convex Optimisation* from Stanford professors who have implemented cvxopt and convex optimisation is a technology especially for L^1 penalties. The general problem above thus has a canonical solution once B is known. Thus research efforts can be localised for determining B and it is known that the quality of recovery of x is extremely good. This reduces moving parts for the problem.

I envision determination of B to have research improvements in the future but I strongly doubt that convex optimisation has to be touched at all.

5. Instrument-Dependent B

The size of B will be roughly 100,000 times the length of mass spec line. It can be determined once per instrument. The recovery of x is relatively fast. There might be various simplifications for samples that do not have 100,000 different possible protein fragments. But that's minor. The important thing is that the quality of B will affect quality of the answer x. For separations mass spectra analysis is line-by-line.

6. Roche And Genentech Considers The Above Favourably

Thanks to the interest of John Stults, who is Director of Analytical Chemistry at Genentech now, Roche and Genentech considered the transformative potential of the above and it is positive. Being twenty first century corporations, they are strongly anti-discriminatory by race and national origin and such and ran into Bill Gates opposition to my success and his use of racial supremacist tools arrayed against myself.

I was watching *Mank* about Herman Mankowiecz who wrote the screenplay of *Citizen Kane* based on William Randolph Hearst. There is a certain gentility of the American tycoons of a bygone era compared to our crass racial tycoon such as Bill Gates today, but I recommend the film to Roche and Genentech executives.

7. Simple View of Proteomic Field

Sometimes a vastly simplified view of the proteomics field will give us some insights since hundreds of thousands of people are involved. What we would like, in dreamland, is to have devices that tell us the list of proteins and their concentrations in biological fluids, serum or other. The ideal instrument we can imagine cartoonishly simply as a miniature device that without troubles quickly tells us the protein concentrations even of complex mixtures as human blood serum.

We don't have such devices. Instead we have separations mass spectrometry and the output are not list of proteins and their concentrations but gigantic two dimensional images of spectra of *fragments of ionised proteins in an unidentifiable complex spectrum.*

What is missing here is an integration of mathematical or analytically sophisticated integration of various contingencies of instrumentation. The instruments are complex with ionisation, separations and mass spectrometry with a variety of features.

What I am proposing is that with the help of sophisticated advances in *convex optimisation* we could have a blind "integration" of two-dimensional mass spectra that is universal in scope, a single canonical method to obtain the biologically most valuable reading of the fluid, just the concentration of intact proteins in the sample.

This is a "post processing" step that at first seems minor but in fact accurate integration is possible for any instrument combination (with effort). Every single bioinformatic and other analytical task is infinitely simpler than currently possible after we have these measurements because the curse of dimensionality and other sample size problems are all instantly alleviated give success in integration. Classical multivariate statistics is suddenly effective again, biomarkers are much closer to the actual biology of the system.

This sort of advance seems a bit unrealistic at first but once people are habituated to using the post-processed output of protein concentrations, I suggest that they will consider the days of 2D mass spectra to be arduous and tedious primitive times as happens with all large technological innovations. The workflow of all proteomic bioinformatics will adapt to the innovation and become substantially more powerful.

8. Significant Application Of Convex Optimisation

There is a great account of history of convex optimisation. Sparsity of solutions is the key feature of L^1 regularised fitting. In basis pursuit methods they are used and I am quite confident that addressing the signal processing of proteomic mass spectra will be one of the most significant nontrivial scientific applications of this body of work. It is appropriate tribute to the efforts in the realm of convex optimisation and billions of people will benefit when proteomics adopts the clear and canonical application of convex optimisation on this problem.

The best book on convex optimisation is by Stephen Boyd and Lieven Vandeberghe published by Cambridge University Press. It is a technology that is well-established. We want to make the use of this technology to proteomics canonical so that the entire field can update its power of analysis of proteome.

9. Emphasis on General Solution For Field of Proteomics

The application of convex optimisation for post-processing of separations mass spectra is not as valuable for ad hoc work of an individual bioinformatics group. It is much more valuable as a sea-change for workflow of the entire field of proteomics. I therefore want to propose that there is a canonical "right way" of doing this with various theoretical justifications with convex optimisations and the entire field of proteomics is better off adopting the method for benefit of all proteomic bioinformatics workflows and that will benefit all subsequent scientific applications of proteomics in the future.

10. Proteomics Will Be Crippled By Curse of Dimensions Without Canonical Post-Processing

From its very beginning statistics of high dimensional data and the curse of dimensionality had plagued proteomics. The sample sizes are not in the range for statistical power for conclusions with separations mass spectra. My appeal for post-processing is absolutely necessary for data for which classical (or even non-classical) multivariate statistical techniques will show some power for results. From the list of protein (neutral) masses and concentrations, which is a much better replacement than blind algorithmic dimension reduction techniques or unreliable peak-picking we will have efficacy for multivariate statistics. There are no clever tricks that can produce a more felicitous situation in proteomics.

11. Not A Life Goal

It is not a life goal of mine to transform the way proteomics is done in the future. Rather it seems simple good sense that having proteins and their concentrations instead of analysing the raw separations mass spectra is a wiser course of action for proteomic analyses in various applications such as biomarker discovery.

12. I Deny Responsibility For Transformation Of Proteomics

I believe it would be a universal good *if* proteomics were to change to conform to the sort of vision I have, but I do not want this challenge to be my burden. I do not know how to produce consensus across many people in biotech who have other opinions and so on.

13. Broad Consensus Is Very Hard

I have been trying to produce consensus among world's scientists to adopt my four-sphere theory for a Scientific Revolution centered around my own work from 2008 to 2012 for many years and I have not succeeded. I don't want any job where I have to succeed in producing grand consensus. I will leave that to more skilled people in rhetoric and have influence enough to succeed.

All I will say is that using my strategy to integrate the separations mass spectra before subsequent analyses is a good idea, very good idea. It will make further analyses infinitely more possible.

I worked for Predicant/Biospect for roughly three years. In this time, analyses without protein list reduction was a constant battle with small samples and curse of dimensionality and impossibly difficult statistical learning problems.

I have a desire to promote not only a Scientific Revolution that is based on my four-sphere theory but also some of my interpretations of poetry of Rainer Maria Rilke. I feel that my interpretation of Duino Elegies have particular novel elements and validity. I introduce 'mythological closure' as a concept to produce the world according to poetic imagination that is totalising and then I interpret the stage of Rilke as all of existence of the subjective reality of the poet. Only from this perspective does the vibrant precision of language of Rilke stand to tell us of how existence is experienced.

14. I Want A Relaxed Job Doing Some Good Science

I have some experience in biotechnology working on problems of signal processing and pattern recognition of mass spectrometry data. I would like to do these things in a relaxed manner without responsibility for grand consensus building that is an impossibly difficult task for me.

15. My Work In The Last 12 Years

From 2008 to 2018 my major work was to work out four-sphere theory that I did without any institutional assistance living in Allen Texas with my aunt. The work was successful [2]. My accomplishments of this period include a refutation of gravitational force and of expansion and big bang which are proven on data.

My four-sphere theory is totally original and is material that is appropriate for a vast scientific revolution of the 1900-1930 paradigm and I deserve, by its merits, to get tenure level rewards at Harvard, Stanford, Oxford and Cambridge. The work was done without employment while I stayed at my aunt's house between 2008-2018. I have further discoveries in human nature and psychology as well.

16. I Will Need More Help From Harvard Board For My Positive Psychology Ideas

I think Harvard Board can do more than give me tenure and funding of order $100 million for quantitative positive psychology and preparation to manage world's life satisfaction. I will need some assistance in getting good infrastructure to recruit talent and manage the organisation. This will be a pioneering effort and will require a great deal of new thinking in incubation.

17. Harvard Board Ought To Confirm In Writing

I am in financial uncertainty in my life, as the options for a proteomics job are extremely uncertain. Harvard Board should evaluate my public work representing the last decade's efford and provide me with tenure (adjunct at Stanford for example) and provide me with funding for a visionary project for quantitative positive psychology and management eventually of world's life satisfaction. It is tremendously difficult to keep focus on work as daily distractions dilute the vision without funding and certainty.

It will be a missed opportunity for the entire world if I am not given the ability to push the project forward as these ideas originate with me.

I am not used to any lavish lifestyle and so the chances for misuse of funds and such will not arise at all. I feel comfortable with the lifestyle of an urban professional and will relocate to San Francisco to do the work. Delays will degrade my ability to keep focus and deliver something acceptable to everyone.

18. Albert Ellis and Aaron T. Beck

In *Learned Optimism* Martin Seligman reveals his erudition by covering the work of Albert Ellis and Aaron T. Beck on Depression. The general idea is to reduce preconceptions of the "shoulds" in people's hopes. I want to point out that the advances are neither well-known outside psychologist circles and they are also not being used on massive scale to promote well-being of people as a life satisfaction. Martin Seligman's general proposal is that optimism enhances quality of life for everyone and there are ways to learn to practice it. My idea of quantitative positive psychology is geared towards managing life satisfaction by disseminating services that are individualised on a massive scale following Martin Seligman's proposal for everyone in a precise way by using communication technology.

I have worked at Lehman Brothers and several startup ventures and believe that I will do a reasonably good job in using funding for a positive psychology/quant venture with some effectiveness. I am seeking order $100 million for a pilot operation. Harvard Board will not regret a grant to me.

In general eight billion people on Earth need some reasonably advanced and effective psychological assistance that is neither based on psychoanalysis whose explanations have not borne sufficient fruit in results, nor on biomedical obsession with bad brain chemistry but scientifically more cogent effort, and this is not to treat pathologies but to maintain good psychologically sound practices involving optimism and gratitude that are religion-independent.

19. My Optimism About The Future

Little by little, through strenuous struggle, I had overthrown the entire theoretical physics apparatus of a century, the 1900-1930 paradigm, by my own efforts, isolated from the central hubs of physics of this age, living on disability money in my aunt's house. I believe that this success shows that I can accomplish a successful flowering of the ideas of quantitative positive psychology and life satisfaction management given the tenure and funding that I seek from Harvard Board. I am optimistic about my proven abilities to overcome challenges of a severe magnitude. I hope this is clear to the Harvard Board.

References

[1] `https://github.com/zulf73/s4sim`
[2] `https://github.com/zulf73/S4TheoryNotes`
[3] `https://github.com/zulf73/moral-nature-virtues`
[4] S. Chen, D. Donoho and M. Saunders, Atomic Decomposition by Basis Pursuit, SIAM Review, 2001, 43 (1), 129–159